(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 June 2003 (05.06.2003)

PCT

(10) International Publication Number
WO 03/046149 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/38441

(22) International Filing Date: 27 November 2002 (27.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/333,966 28 November 2001 (28.11.2001) US

(71) Applicant (*for all designated States except US*): **MJ BIOWORKS INCORPORATED** [US/US]; 7000 Shoreline Court, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **WANG, Yan** [CN/US]; 1370 11th Avenue, #6, San Francisco, CA 94122 (US). **VANDER HORN, Peter** [US/US]; 130 Trimaran Court, Foster City, CA 94404 (US). **XI, Lei** [US/US]; 134 E. Court Lane, Foster City, CA 94404 (US).

(74) Agents: **LOCKYER, Jean, M.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

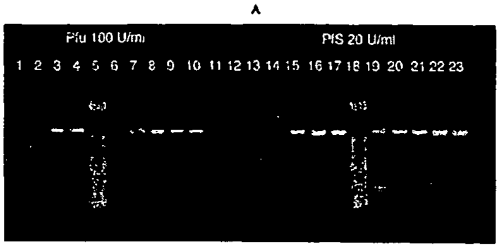
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),

[Continued on next page]

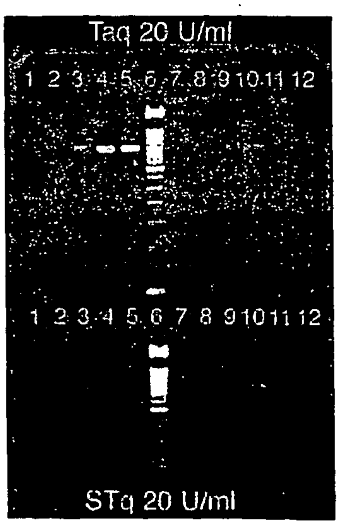
(54) Title: METHODS OF USING IMPROVED POLYMERASES

(57) Abstract: This invention provides for methods of sequencing and performing polymerase reactions using an improved generation of nucleic acid polymerases. The improvement is the fusion of a sequence-non-specific nucleic-acid-binding domain to the enzyme in a manner that enhances the processivity of the polymerase.

A



B



WO 03/046149 A2

European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS OF USING IMPROVED POLYMERASES

[0001] This application claims the benefit of U.S. provisional application no. 60//333,966, filed November 28, 2001, which is incorporated by reference herein.

FIELD OF THE INVENTION

[0002] This invention provides more efficient methods of performing polymerase reactions. The methods employ an improved generation of nucleic acid polymerases. The improvement is the joining sequence-non-specific nucleic-acid-binding domain to the enzyme in a manner that enhances the ability of the enzyme to bind and catalytically modify the nucleic acid.

BACKGROUND OF THE INVENTION

[0003] The processivity of a polymerase, *i.e.*, the amount of product generated by the enzyme per binding event, can be enhanced by increasing the stability of the modifying enzyme/nucleic acid complex. The current invention now provides enhanced polymerase assays that employ novel modifying enzymes in which the double-stranded conformation of the nucleic acid is stabilized and the processivity of the enzyme increased by joining a sequence-non-specific double-stranded nucleic acid binding domain to the enzyme, or its catalytic domain which are disclosed *e.g.,* in co-pending U.S. Application No. 09/870,353 and WO01/92501. The modifying proteins that are processive in nature exhibit increased processivity when joined to a binding domain compared to the enzyme alone.

[0004] There is a need to enhance polymerase reactions in many applications. For example, SYBR Green I (Molecular Probes, Eugene, OR; US Patents 5,436,134 and 5,658,751), a fluorescent dye that is specific for dsDNA detection, is widely used in real-time PCR reactions to monitor the generation of dsDNA through each cycle of amplification. However, the addition of SYBR Green I inhibits the activity of DNA polymerases used in PCR. Similarly, it is often desirable to use PCR for the analysis of crude or "dirty" nucleic acid samples. For example, colony PCR is a useful technique in which small samples of single bacterial colonies are lysed and added directly to PCR reactions for the purpose of screening colonies for particular DNA sequences. However, colony PCR has a high failure rate, because of residual contaminants from the colony. Thus, polymerases that are resistant

to such inhibitors, *e.g.*, fluorescent dyes and impurities present in the cell extracts, are needed in order to obtain more efficient polymerase reactions, *e.g.*, PCR.

[0005] There is also a need to improve sequencing reactions. Polymerases currently employed in sequencing reactions, *e.g.*, cycle sequencing, are often inefficient. For example, cycle sequencing is often performed with poorly-processive enzymes. Often, the enzymes used are ΔTaq derivatives, which have Taq polymerase's 5'-3' nuclease domain removed, and have a processivity of about 2 bases. Also, in the case of dye terminator-sequencing, dITP is used in place of dGTP, which causes polymerase pausing and dissociation at G nucleotides. These enzymes therefore produce a large number of sequence products that are improperly terminated. These stops compete with, and negatively effect, the production of properly terminated sequence products. Furthermore, if a polymerase dissociates during primer extension of a template containing a repeat unit (*e.g.*, a triplet repeat) or secondary structure (*e.g.*, a stem and loop), the 3' end can denature and reanneal so as to prime at a different location on the template — for example, in the case of a repeat, the reannealing could occur at a different repeat; or in the case of secondary structure, improper reannealing could delete out a section of the template. Thus, dissociation of the polymerase during sequencing can cause a problem in efficiently obtaining reliable sequencing information.

[0006] The current invention addresses both of these needs, *i.e.*, the need for enhancing polymerase reactions performed in the presence of inhibitors and the need for enhancing processivity in DNA sequencing applications). The current invention provides such enhanced, or improved, polymerase reactions. The improvement is the use of a polymerase that has increased processivity due to the presence of a sequence-non-specific nucleic-acid-binding domain that is joined to the polymerase.

BRIEF SUMMARY OF THE INVENTION

[0007] The present invention provides methods of performing more efficient polymerase reactions using a polymerase protein comprising a polymerase domain joined to a sequence-non-specific double-stranded nucleic acid binding domain. Typically the presence of the sequence non-specific double-stranded nucleic acid binding domain enhances the processivity of the polymerase compared to an identical protein not having a sequence-non-specific nucleic acid binding domain joined thereto.

[0008] The polymerase domain can be thermally stable, *e.g.*, a *Thermus* polymerase domain such as a ΔTaq polymerase domain, or a *Pyrococcus* polymerase domain.

[0009] In one embodiment the sequence-non-specific nucleic-acid-binding domain specifically binds to polyclonal antibodies generated against either Sac7d or Sso7d. Alternatively, the sequence-non-specific nucleic-acid-binding domain contains a 50 amino acid subsequence containing 50% amino acid similarity to Sso7d. Typically, the sequence-non-specific nucleic-acid-binding domain is Sso7d or specifically binds to polyclonal antibodies generated against Sso7d.

[0010] The polymerase reaction can be performed on a target nucleic acid that is present in a crude preparation of a sample. In another embodiment, the polymerase reaction is performed in the presence of a molecule that typically inhibits polymerases, *e.g.* fluorescent dyes such as SYBR Green I. Further, the polymerase may be used in cycle sequencing reactions to obtain longer sequences, *e.g.*, through regions of secondary structure that prevent sequencing using unmodified polymerases.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] Figures 1A and 1B show the results of a PCR reaction performed in the presence of contaminants using an improved polymerase.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0012] "Archaeal small basic DNA-binding protein" refers to protein of between 50-75 amino acids having either 50% homology to a natural Archaeal small basic DNA-binding protein such as Sso-7d from *Sulfolobus sulfataricus* or binds to antibodies generated against a native Archaeal small basic DNA-binding protein.

[0013] "Domain" refers to a unit of a protein or protein complex, comprising a polypeptide subsequence, a complete polypeptide sequence, or a plurality of polypeptide sequences where that unit has a defined function. The function is understood to be broadly defined and can be ligand binding, catalytic activity or can have a stabilizing effect on the structure of the protein.

[0014] "Efficiency" in the context of a nucleic acid modifying enzyme of this invention refers to the ability of the enzyme to perform its catalytic function under specific reaction conditions. Typically, "efficiency" as defined herein is indicated by the amount of product generated under given reaction conditions.

[0015] "Enhances" in the context of an enzyme refers to improving the activity of the enzyme, *i.e.*, increasing the amount of product per unit enzyme per unit time.

[0016] "Fused" refers to linkage by covalent bonding.

[0017] "Heterologous", when used with reference to portions of a protein, indicates that the protein comprises two or more domains that are not found in the same relationship to each other in nature. Such a protein, *e.g.*, a fusion protein, contains two or more domains from unrelated proteins arranged to make a new functional protein.

[0018] "Join" refers to any method known in the art for functionally connecting protein domains, including without limitation recombinant fusion with or without intervening domains, intein-mediated fusion, non-covalent association, and covalent bonding, including disulfide bonding; hydrogen bonding; electrostatic bonding; and conformational bonding, *e.g.*, antibody-antigen, and biotin-avidin associations.

[0019] "Nucleic-acid-modifying enzyme" refers to an enzyme that covalently alters a nucleic acid.

[0020] "Polymerase" refers to an enzyme that performs template-directed synthesis of polynucleotides. The term, as used herein, also refers to a domain of the polymerase that has catalytic activity.

[0021] "Error-correcting activity" of a polymerase or polymerase domain refers to the 3' to 5' exonuclease proofreading activity of a template-specific nucleic acid polymerase whereby nucleotides that do not form Watson-Crick base pairs with the template are removed from the 3' end of an oligonucleotide, *i.e.*, a strand being synthesized from a template, in a sequential manner. Examples of polymerases that have error-correcting activity include polymerases from *Pryococcus furiosus, Thermococcus litoralis,* and *Thermotoga maritima*.

[0022] Processivity refers to the ability of a nucleic acid modifying enzyme to remain bound to the template or substrate and perform multiple modification reactions. Processivity is measured by the number of catalytic events that take place per binding event.

[0023] "Sequence-non-specific nucleic-acid-binding domain" refers to a protein domain which binds with significant affinity to a nucleic acid, for which there is no known nucleic acid which binds to the protein domain with more than 100-fold more affinity than another nucleic acid with the same nucleotide composition but a different nucleotide sequence.

[0024] "Thermally stable polymerase" as used herein refers to any enzyme that catalyzes polynucleotide synthesis by addition of nucleotide units to a nucleotide chain using DNA or RNA as a template and has an optimal activity at a temperature above 45°C.

[0025] "*Thermus* polymerase" refers to a family A DNA polymerase isolated from any *Thermus* species, including without limitation *Thermus aquaticus, Thermus brockianus, and Thermus thermophilus*; any recombinant enzymes deriving from *Thermus* species, and any

functional derivatives thereof, whether derived by genetic modification or chemical modification or other methods known in the art.

[0026] The term "amplification reaction" refers to any *in vitro* means for multiplying the copies of a target sequence of nucleic acid. Such methods include but are not limited to polymerase chain reaction (PCR), DNA ligase reaction (*see* U.S. Patents 4,683,195 and 4,683,202; *PCR Protocols: A Guide to Methods and Applications* (Innis *et al.*, eds, 1990)), (LCR), QBeta RNA replicase, and RNA transcription-based (such as TAS and 3SR) amplification reactions as well as others known to those of skill in the art.

[0027] "Amplifying" refers to a step of submitting a solution to conditions sufficient to allow for amplification of a polynucleotide if all of the components of the reaction are intact. Components of an amplification reaction include, e.g., primers, a polynucleotide template, polymerase, nucleotides, and the like. The term "amplifying" typically refers to an "exponential" increase in target nucleic acid. However, "amplifying" as used herein can also refer to linear increases in the numbers of a select target sequence of nucleic acid.

[0028] The term "amplification reaction mixture" refers to an aqueous solution comprising the various reagents used to amplify a target nucleic acid. These include enzymes, aqueous buffers, salts, amplification primers, target nucleic acid, and nucleoside triphosphates. Depending upon the context, the mixture can be either a complete or incomplete amplification reaction mixture

[0029] "Polymerase chain reaction" or "PCR" refers to a method whereby a specific segment or subsequence of a target double-stranded DNA, is amplified in a geometric progression. PCR is well known to those of skill in the art; *see, e.g.*, U.S. Patents 4,683,195 and 4,683,202; and *PCR Protocols: A Guide to Methods and Applications*, Innis *et al.*, eds, 1990. Exemplary PCR reaction conditions typically comprise either two or three step cycles. Two step cycles have a denaturation step followed by a hybridization/elongation step. Three step cycles comprise a denaturation step followed by a hybridization step followed by a separate elongation step.

[0030] "Long PCR" refers to the amplification of a DNA fragment of 5 kb or longer in length. Long PCR is typically performed using specially-adapted polymerases or polymerase mixtures (*see, e.g.*, U.S. Patent Nos. 5, 436,149 and 5,512,462) that are distinct from the polymerases conventionally used to amplify shorter products.

[0031] A "primer" refers to a polynucleotide sequence that hybridizes to a sequence on a target nucleic acid and serves as a point of initiation of nucleic acid synthesis. Primers can be of a variety of lengths and are often less than 50 nucleotides in length, for example 12-30

nucleotides, in length. The length and sequences of primers for use in PCR can be designed based on principles known to those of skill in the art, *see, e.g.*, Innis et al., *supra.*

[0032] A temperature profile refers to the temperature and lengths of time of the denaturation, annealing and/or extension steps of a PCR or cycle sequencing reaction. A temperature profile for a PCR or cycle sequencing reaction typically consists of 10 to 60 repetitions of similar or identical shorter temperature profiles; each of these shorter profiles may typically define a two step or three-step cycle. Selection of a temperature profile is based on various considerations known to those of skill in the art, *see, e.g.*, Innis *et al.*, *supra.* In a long PCR reaction as described herein, the extension time required to obtain an amplification product of 5 kb or greater in length is reduced compared to conventional polymerase mixtures.

[0033] PCR "sensitivity" refers to the ability to amplify a target nucleic acid that is present in low copy number. "Low copy number" refers to 10^5, often 10^4, 10^3, 10^2, or fewer, copies of the target sequence in the nucleic acid sample to be amplified.

[0034] A "template" refers to a double stranded polynucleotide sequence that comprises the polynucleotide to be amplified, flanked by primer hybridization sites. Thus, a "target template" comprises the target polynucleotide sequence flanked by hybridization sites for a 5' primer and a 3' primer.

[0035] An "improved polymerase" includes a sequence-non-specific double-stranded DNA binding domain joined to the polymerase or polymerase domain. An "unimproved polymerase" is a polymerase that does not have a sequence-non-specific double-stranded DNA binding domain.

Introduction

[0036] The current invention provides methods of performing polymerase reactions using improved polymerases. These polymerase reactions are typically more efficient and yield more product than traditional polymerases. These improved polymerases contain a polymerase domain with a binding domain joined to it. While the prior art taught that nucleic acid binding proteins can increase the binding affinity of enzymes to nucleic acid, the group of binding proteins having the ability to enhance the processive nature of the enzymes is of particular value. Not to be bound by theory, binding domains of the invention typically dissociate from double-stranded nucleic acid at a very slow rate. Thus, they increase the processivity and/or efficiency of a modifying enzyme to which they are joined by stabilizing the enzyme-nucleic acid complex. Accordingly, this invention results from the discovery that

DNA-binding domains can stabilize the double-stranded conformation of a nucleic acid and increase the efficiency of a catalytic domain that requires a double-stranded substrate. Described herein are examples and simple assays to readily determine the improvement to the catalytic and/or processive nature of catalytic nucleic acid modifying enzymes, *e.g.*, polymerases.

Polymerase Domains.

[0037] DNA polymerases are well-known to those skilled in the art. These include both DNA-dependent polymerases and RNA-dependent polymerases such as reverse transcriptase. At least five families of DNA-dependent DNA polymerases are known, although most fall into families A, B and C. There is little or no structural or sequence similarity among the various families. Most family A polymerases are single chain proteins that can contain multiple enzymatic functions including polymerase, 3' to 5' exonuclease activity and 5' to 3' exonuclease activity. Family B polymerases typically have a single catalytic domain with polymerase and 3' to 5' exonuclease activity, as well as accessory factors. Family C polymerases are typically multi-subunit proteins with polymerizing and 3' to 5' exonuclease activity. In *E. coli*, three types of DNA polymerases have been found, DNA polymerases I (family A), II (family B), and III (family C). In eukaryotic cells, three different family B polymerases, DNA polymerases α, δ, and ε, are implicated in nuclear replication, and a family A polymerase, polymerase γ, is used for mitochondrial DNA replication. Other types of DNA polymerases include phage polymerases.

[0038] Similarly, RNA polymerases typically include eukaryotic RNA polymerases I, II, and III, and bacterial RNA polymerases as well as phage and viral polymerases. RNA polymerases can be DNA-dependent and RNA-dependent.

[0039] In one embodiment, polymerase domains that have an error-correcting activity are used as the catalytic domain of the improved polymerases described herein. These polymerases can be used to obtain long, *i.e.*, 5 kb, often 10 kb, or greater in length, PCR products. "Long PCR" using these improved polymerases can be performed using extension times that are reduced compared to prior art "long PCR" polymerase and/or polymerase mixtures. Extension times of less than 30 seconds per kb, often 15 seconds per kb, can be used to amplify long products in PCR reactions using the improved polymerases. Furthermore, these modified polymerases also exhibit increased sensitivity.

[0040] Prior-art non-error-correcting polymerases such as Taq polymerase are capable of amplifying DNA from very small input copy concentrations, such as, in the extreme, 10 copies per ml. However, because of the low fidelity of such polymerases, products cloned from such amplifications are likely to contain introduced mutations.

[0041] Prior-art error-correcting polymerases such as Pfu copy DNA with higher fidelity than Taq, but are not capable of amplifying DNA from small input copy concentrations. The hybrid error-correcting polymerases of the invention exhibit much higher processivity while retaining error-correcting activity and thereby provide both sensitivity and fidelity in amplification reactions.

[0042] The activity of a polymerase can be measured using assays well known to those of skill in the art. For example, a processive enzymatic activity, such as a polymerase activity, can be measured by determining the amount of nucleic acid synthesized in a reaction, such as a polymerase chain reaction. In determining the relative efficiency of the enzyme, the amount of product obtained with a polymerase containing a sequence-non-specific double-stranded DNA binding domain can then be compared to the amount of product obtained with the normal polymerase enzyme, which will be described in more detail below and in the Examples.

[0043] A polymerase domain suitable for use in the invention can be the enzyme itself or the catalytic domain, *e.g.*, Taq polymerase or a domain of Taq with polymerase activity. The catalytic domain may include additional amino acids and/or may be a variant that contains amino acid substitutions, deletions or additions, but still retains enzymatic activity.

Sequence-Non-Specific Nucleic-Acid-Binding Domain

[0044] A double-stranded sequence-non-specific nucleic acid binding domain is a protein or defined region of a protein that binds to double-stranded nucleic acid in a sequence-independent manner, *i.e.*, binding does not exhibit a gross preference for a particular sequence. Typically, double-stranded nucleic acid binding proteins exhibit a 10-fold or higher affinity for double-stranded versus single-stranded nucleic acids. The double-stranded nucleic acid binding proteins in particular embodiments of the invention are preferably thermostable. Examples of such proteins include, but are not limited to, the Archaeal small basic DNA binding proteins Sac7d and Sso7d (*see, e.g.*, Choli *et al.*, *Biochimica* et *Biophysica Acta* 950:193-203, 1988; Baumann *et al.*, *Structural Biol.* 1:808-819, 1994; and Gao *et al*, *Nature Struc. Biol.* 5:782-786, 1998), Archael HMf-like proteins (*see, e.g.*, Starich *et al.*, *J. Molec. Biol.* 255:187-203, 1996; Sandman *et al.*, *Gene* 150:207-208, 1994), and

PCNA homologs (*see, e.g.*, Cann *et al.*, *J. Bacteriology* 181:6591-6599, 1999; Shamoo and Steitz, *Cell*:99, 155-166, 1999; De Felice *et al.*, *J. Molec. Biol.* 291, 47-57, 1999; and Zhang *et al.*, *Biochemistry* 34:10703-10712, 1995).

Sso7d and Sac7d

[0045] Sso7d and Sac7d are small (about 7,000 kd MW), basic chromosomal proteins from the hyperthermophilic archaeabacteria *Sulfolobus solfataricus* and *S. acidocaldarius*, respectively. These proteins are lysine-rich and have high thermal, acid and chemical stability. They bind DNA in a sequence-independent manner and when bound, increase the T_M of DNA by up to 40° C under some conditions (McAfee *et al.*, *Biochemistry* 34:10063-10077, 1995). These proteins and their homologs are typically believed to be involved in packaging genomic DNA and stabilizing genomic DNA at elevated temperatures.

HMF-like proteins

[0046] The HMf-like proteins are archaeal histones that share homology both in amino acid sequences and in structure with eukaryotic H4 histones, which are thought to interact directly with DNA. The HMf family of proteins form stable dimers in solution, and several HMf homologs have been identified from thermostable species (*e.g.*, *Methanothermus fervidus* and *Pyrococcus* strain GB-3a). The HMf family of proteins, once joined to Taq DNA polymerase or any DNA modifying enzyme with a low intrinsic processivity, can enhance the ability of the enzyme to slide along the DNA substrate and thus increase its processivity. For example, the dimeric HMf-like protein can be covalently linked to the N terminus of Taq DNA polymerase, *e.g.*, via chemical modification, and thus improve the processivity of the polymerase.

PCNA homologs

[0047] Many but not all family B DNA polymerases interact with accessory proteins to achieve highly processive DNA synthesis. A particularly important class of accessory proteins is referred to as the sliding clamp. Several characterized sliding clamps exist as trimers in solution, and can form a ring-like structure with a central passage capable of accommodating double-stranded DNA. The sliding clamp forms specific interactions with the amino acids located at the C terminus of particular DNA polymerases, and tethers those polymerases to the DNA template during replication. The sliding clamp in eukarya is referred to as the proliferating cell nuclear antigen (PCNA), while similar proteins in other

domains are often referred to as PCNA homologs. These homologs have marked structural similarity but limited sequence similarity.

[0048] Recently, PCNA homologs have been identified from thermophilic Archaea (*e.g.*, *Sulfalobus sofataricus*, *Pyroccocus furiosus*, etc.). Some family B polymerases in Archaea have a C terminus containing a consensus PCNA-interacting amino acid sequence and are capable of using a PCNA homolog as a processivity factor (*see, e.g.*, Cann *et al.*, *J. Bacteriol.* 181:6591-6599, 1999 and De Felice *et al.*, *J. Mol. Biol.* 291:47-57, 1999). These PCNA homologs are useful sequence-non-specific double-stranded DNA binding domains for the invention. For example, a consensus PCNA-interacting sequence can be joined to a polymerase that does not naturally interact with a PCNA homolog, thereby allowing a PCNA homolog to serve as a processivity factor for the polymerase. By way of illustration, the PCNA-interacting sequence from *Pyrococcus furiosus* PolII (a heterodimeric DNA polymerase containing two family B-like polypeptides) can be covalently joined to *Pyrococcus furiosus* PolI (a monomeric family B polymerase that does not normally interact with a PCNA homolog). The resulting fusion protein can then be allowed to associate non-covalently with the *Pyrococcus furiosus* PCNA homolog to generate a novel heterologous protein with increased processivity relative to the unmodified *Pyrococcus furiosus* PolI.

Other sequence-nonspecific double-stranded nucleic acid binding domains

[0049] Additional nucleic acid binding domains suitable for use in the invention can be identified by homology with known sequence non-specific double-stranded DNA binding proteins and/or by antibody crossreactivity, or may be found by means of a biochemical assay. These methods are described, *e.g.*, in WO01/92501. Further, methods of joining the polymerase to the sequence non-specific double-stranded DNA binding protein and methods of expressing recombinant polymerases and polymerase fusion proteins are also described (*see, e.g.*, WO01/92501).

Assays To Determine Improved Activity of Polymerase Domains.

[0050] Activity of the polymerase domain can be measured using a variety of assays that can be used to compare processivity or modification activity of a modifying protein domain joined to a binding domain compared to the protein by itself. Improvement in activity includes both increased processivity and increased efficiency.

[0051] Polymerase processivity can be measured in variety of methods known to those of ordinary skill in the art. Polymerase processivity is generally defined as the number of

nucleotides incorporated during a single binding event of a modifying enzyme to a primed template.

[0052] For example, a 5' FAM-labeled primer is annealed to circular or linearized ssM13mp18 DNA to form a primed template. In measuring processivity, the primed template usually is present in significant molar excess to the enzyme or catalytic domain to be assayed so that the chance of any primed template being extended more than once by the polymerase is minimized. The primed template is therefore mixed with the polymerase catalytic domain to be assayed at a ratio such as approximately 4000:1 (primed DNA:DNA polymerase) in the presence of buffer and dNTPs. $MgCl_2$ is added to initiate DNA synthesis. Samples are quenched at various times after initiation, and analyzed on a sequencing gel. At a polymerase concentration where the median product length does not change with time or polymerase concentration, the length corresponds to the processivity of the enzyme. The processivity of a protein of the invention, *i.e.*, a protein that contains a sequence non-specific double-stranded nucleic acid binding domain fused to the catalytic domain of a processive nucleic acid modifying enzyme such as a polymerase, is then compared to the processivity of the enzyme without the binding domain.

[0053] Enhanced efficiency can also be demonstrated by measuring the increased ability of an enzyme to produce product. Such an analysis measures the stability of the double-stranded nucleic acid duplex indirectly by determining the amount of product obtained in a reaction. For example, a PCR assay can be used to measure the amount of PCR product obtained with a short, *e.g.*, 12 nucleotide in length, primer annealed at an elevated temperature, *e.g.*, 50°C. In this analysis, enhanced efficiency is shown by the ability of a polymerase such as a Taq polymerase to produce more product in a PCR reaction using the 12 nucleotide primer annealed at 50°C when it is joined to a sequence-non-specific double-stranded nucleic-acid-binding domain of the invention, *e.g.*, Sso7d, than Taq polymerase does alone. In contrast, a binding tract that is a series of charged residues, *e.g.* lysines, does not enhance processivity when joined to a polymerase.

[0054] Assays such as salt sensitivity can also be used to demonstrate improvement in efficiency of a processive nucleic acid modifying enzyme of the invention. A modifying enzyme, or the catalytic domain, when fused to a sequence non-specific double-stranded nucleic acid binding domain of the invention exhibits increased tolerance to high salt concentrations, *i.e.*, a processive enzyme with increased processivity can produce more product in higher salt concentrations . For example, a PCR analysis can be performed to determine the amount of product obtained in a reaction using a fusion Taq polymerase (e.g.,

Sso7d fused to Taq polymerase) compared to Taq polymerase in reaction conditions with high salt, *e.g.*, 80 mM.

[0055] Other methods of assessing enhanced efficiency of the improved polymerases of the invention can be determined by those of ordinary skill in the art using standard assays of the enzymatic activity of a given modification enzyme.

[0056]

Uses of improved polymerases

[0057] The invention provides improved methods of performing polymerase reactions. In one embodiment, the invention provides a method of performing a polymerase reaction in the presence of a fluorescent dye. A number of fluorescent dyes that are commonly used in reactions such as real-time PCR, have an inhibitory activity on polymerases have been typically used in PCR, *e.g.*, Taq polymerase. For example, SYBR Green I (Molecular Probes, Eugene, OR; US Patents 5,436,134 and 5,658,751), is a fluorescent dye that is specific for dsDNA detection, and is widely used in real-time PCR reactions to monitor the generation of dsDNA through each cycle of amplification. Use of dyes to monitor amplification is described in US Patents 5,994,056 and 6,171,785 and use of SYBR Green I for this purpose is described in Morrison *et al.*, *Biotechniques* 24:954-962 (1998).

[0058] It has been observed that the addition of SYBR Green I inhibits the activity of DNA polymerases used in PCR, possibly through interfering with the binding of the polymerase to the primer-template. Additives such as DMSO are therefore often required to reduce the inhibitory effect of the dye. However, DMSO can reduce the storage stability of the enzyme and can inhibit polymerases. The current invention provides a method of performing polymerase reactions in the presence of a fluorescent dye that uses the improved polymerases described herein, which are not as sensitive to the fluorescent dye, *i.e.*, are not inhibited to the same extent, as an unimproved polymerase.

[0059] The ability of a polymerase to perform in the presence of a dye that exhibits altered fluorescence emissions when bound to double-stranded DNA can be measured using well known polymerase assays, such as those described herein. Typically, a fluorescent dye reduces the activity of an unimproved polymerase by 25%, often 50%, 75%, or more. Polymerase activity can be assayed using the methods described herein.

[0060] The ability of an improved polymerase to perform a PCR reaction in the presence of a fluorescent dye, *e.g.*, SYBR Green I, can also be compared to the ability of the unimproved polymerase to perform in an otherwise identical PCR reaction. The comparison can be made using a values such as the cycle threshold (C_t) value, which represents the number of cycles

required to generate a detectable amount of DNA. An efficient polymerase may be able to produce a detectable amount of DNA in a smaller number of cycles by more closely approaching the theoretical maximum amplification efficiency of PCR. Accordingly,, a lower C_t value reflects a greater amplification efficiency for the enzyme. The improved enzymes exhibit 2x, often 5x, or greater activity in the presence of a fluorescent dye when compared to the unimproved enzyme.

[0061] In typical embodiments, the polymerase reaction is performed in the presence of a fluorescent dye such as SYBR Green I or Pico Green I (Molecular Probes, Eugene, OR;). These dyes are unsymmetrical cyanine dyes containing a defined substituent on the pyridinium or quinolinium ring system or a substituent immediately adjacent to the nitrogen atom of the pyridinium or quinolinium ring. These and other members of the same class of dyes are described, *e.g.*, in US Patent Nos. 5,436,134 and 5,658,751. SYBR Green I, for example, binds specifically to dsDNA with a dissociation constant in the sub-micromolar range. Upon binding, it has a large increase in its quantum yield and therefore a large increase in fluorescence.

[0062] In other embodiments, the polymerase reactions of the invention can be performed in the presence of other fluorescent compounds that typically inhibit polymerases, such as other fluorescent dyes, *e.g.*, propidium iodide, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, mithramycin, ruthenium polypyridyls, and anthramycin, which also exhibit altered fluorescence emissions when bound to double-stranded DNA. Improved polymerases can be tested for resistance to other dyes using methodology well known in the arts and described herein (*see, e.g.*, Example 6).

[0063] In another embodiment, the invention provides method of performing a polymerase reaction, *e.g.*, a PCR reaction, in the presence of contaminants such as those present when using a crude nucleic acid sample. Inhibitors of polymerase activity are often present in crude nucleic acid sample preparations, thus presenting difficulties in using such preparations in polymerase reactions such as PCR or nucleic acid sequence. The improved enzymes are more tolerant to such contaminants. Accordingly, the improved enzymes offer advantages over standard enzymes when performing polymerase reactions, *e.g.* PCR, using crude nucleic acid preparations. These preparations can be from a variety of sources, including cells such as bacterial cells, plant cells, and various other cell types.

[0064] A crude nucleic acid sample typically includes contaminants that originate from the nucleic acid source or from a previous chemical or molecular biological manipulation. The

improved polymerases are less sensitive to the presence of such contaminants. As noted above, polymerase activity assays can be performed using methods described herein. An improved polymerase typically exhibits 2x, 5x, 10x, or greater activity relative to the unimproved polymerase when assayed in the presence of contaminants in an otherwise identical polymerase activity assay or PCR. An exemplary analysis of polymerase activity in crude preparations is provided in Example 7. Crude preparations typically are not processed through repeated rounds of purification and are typically less than 98% pure, often less than 95% pure.

[0065] The modified polymerase enzymes are also more resistant to common additives for troublesome PCR reactions such as Betaine, DMSO, as well as resistant to salt, *e.g.*, KCl, etc. The improved polymerase typically exhibits 2x, 5x, 10x or greater activity relative to the unimproved polymerase in the presence of such agents.

[0066] Improved polymerases can also be used in nucleic acid sequencing reactions. These reactions are well known to those of skill in the art (*see, e.g.*, Sambrook and Russell, *Molecular Cloning, A Laboratory Manual* 3rd. 2001, Cold Spring Harbor Laboratory Press).

[0067] Improved polymerases are particular advantageous when used in sequencing reactions, in particular sequencing reactions that use thermostable polymerases, such as cycle sequencing reactions. Cycle sequencing refers to a linear amplification DNA sequencing technique in which a single primer is used to generate labeled terminated fragments by the Sanger dideoxy chain termination method. Thermostable polymerase enzymes are employed in such reactions.

[0068] Thermostable polymerases such as Taq or Pfu catalyze the incorporation of ddNTPs at a rate that is at least two orders of magnitude slower than the rate of incorporation of dNTPs. In addition, the efficiency of incorporation of a ddNTP at a particular site is affected by the local sequence of the template DNA. Modified version of polymerases that lack 5' to 3' exonuclease activity and catalyze incorporation of ddNTPs with high efficiency have been developed; however, their processivity is often poor. For example, thermostable enzymes are such as ΔTaq derivatives, which have Taq polymerase's 5'-3' nuclease domain removed, have a processivity of about 2 bases. Also, in the case of dye terminator-sequencing, dITP is used in place of dGTP, which causes polymerase pausing and dissociation at G nucleotides. These enzymes therefore produce a large number of sequence products that are improperly terminated. Furthermore, if a polymerase dissociates during primer extension of a template containing a repeat unit (*e.g.*, a triplet repeat) or secondary structure (*e.g.*, a stem and loop) such that the strand is not completed during a particular PCR cycle, the 3' end can denature

and reanneal during a subsequent PCR cycle so as to prime at a different location on the template — for example, in the case of a repeat, the reannealing could occur at a different repeat; or in the case of secondary structure, improper reannealing could delete out a section of the template. Thus, dissociation of the polymerase is also a problem.

[0069] The use of improved polymerases as described herein can provide enhanced sequencing reactions, *e.g.*, cycle sequencing reactions, in which there are fewer improper terminations and fewer dissociation events. This provides longer sequence reads, i.e., the number of nucleotides for which the sequence can be determined, that contain fewer ambiguities compared to reaction performed with unimproved enzymes.

[0070] The polymerases are typically modified to substitute a Y residue for an F residue (US Patent No. 5,614,365).

[0071] All publications, patents, and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

[0072] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

[0073] The following examples are provided by way of illustration only and not by way of limitation. Those of skill will readily recognize a variety of non-critical parameters that could be changed or modified to yield essentially similar results.

Example 1. Construction of fusion proteins.

Construction of Sso7d-ΔTaq fusion.

[0074] The following example illustrates the construction of a polymerase protein possessing enhanced processivity, in which the sequence-non-specific double-stranded nucleic acid binding protein Sso7d is fused to the *Thermus aquaticus* PolI DNA polymerase (a family A polymerase known as Taq DNA polymerase) that is deleted at the N terminus by 289 amino acids (ΔTaq).

[0075] Based on the published amino acid sequence of Sso7d, seven oligonucleotides were used in constructing a synthetic gene encoding Sso7d. The oligonucleotides were annealed and ligated using T4 DNA ligase. The final ligated product was used as the template in a PCR reaction using two terminal oligonucleotides as primers to amplify the full-length gene. By design, the resulting PCR fragment contains a unique EcoRI site at the 5' terminus, and a unique BstXI site at the 3' terminus. In addition to encoding the Sso7d protein, the above PCR fragment also encodes a peptide linker with the amino acid sequence of Gly-Gly-Val-Thr positioned at the C terminus of the Sso7d protein. The synthetic gene of Sso7d has the DNA sequence shown in SEQ ID NO:1, and it encodes a polypeptide with the amino acid sequence shown in SEQ ID NO:2.

[0076] The synthetic gene encoding Sso7d was then used to generate a fusion protein in which Sso7d replaces the first 289 amino acid of Taq. The fragment encoding Sso7d was subcloned into a plasmid encoding Taq polymerase to generate the fusion protein, as follows. Briefly, the DNA fragment containing the synthetic Sso7d gene was digested with restriction endonucleases EcoRI and BstXI, and ligated into the corresponding sites of a plasmid encoding Taq. As the result, the region that encodes the first 289 amino acid of Taq is replaced by the synthetic gene of Sso7d. This plasmid (pYW1) allows the expression of a single polypeptide containing Sso7d fused to the N terminus of ΔTaq via a synthetic linker composed of Gly-Gly-Val-Thr. The DNA sequence encoding the fusion protein (Sso7d-ΔTaq) and the amino acid sequence of the protein are shown in SEQ ID NOs:3 and 4, respectively.

Construction of Sso7d-Taq fusion.

[0077] An Sso7d/full-length Taq fusion protein was also constructed. Briefly, a 1 kb PCR fragment encoding the first 336 amino acids of Taq polymerase was generated using two primers. The 5' primer introduces a SpeI site into the 5' terminus of the PCR fragment, and the 3' primer hybridizes to nucleotides 1008-1026 of the Taq gene. The fragment was digested with SpeI and BstXI, releasing a 0.9 kb fragment encoding the first 289 amino acids of Taq polymerase. The 0.9 kb fragment was ligated into plasmid pYW1 at the SpeI (located in the region encoding the linker) and BstXI sites. The resulting plasmid (pYW2) allows the expression of a single polypeptide containing the Sso7d protein fused to the N terminus of the full length Taq DNA polymerase via a linker composed of Gly-Gly-Val-Thr, the same as in

Sso7d-ΔTaq. The DNA sequence encoding the Sso7d-Taq fusion protein and the amino acid sequence of the protein are shown in SEQ ID. NO.5 and NO.6, respectively.

Construction of Pfu-Sso7d fusion.

[0078] A third fusion protein was created, joining Sso7d to the C terminus of *Pyrococcus furiosus* DNA polI (a family B DNA polymerase known as Pfu). A pET-based plasmid carrying the Pfu DNA polymerase gene was modified so that a unique KpnI site and a unique SpeI site are introduced at the 3' end of the Pfu gene before the stop codon. The resulting plasmid (pPFKS) expresses a Pfu polymerase with three additional amino acids (Gly-Thr-His) at its C terminus.

[0079] Two primers were used to PCR amplify the synthetic Sso7d gene described above to introduce a Kpn I site and a NheI site flanking the Sso7d gene. The 5' primer also introduced six additional amino acids (Gly-Thr-Gly-Gly-Gly-Gly), which serve as a linker, at the N terminus of the Sso7d protein. Upon digestion with KpnI and NheI, the PCR fragment was ligated into pPFKS at the corresponding sites. The resulting plasmid (pPFS) allows the expression of a single polypeptide containing Sso7d protein fused to the C terminus of the Pfu polymerase via a peptide linker (Gly-Thr-Gly-Gly-Gly-Gly). The DNA sequence encoding the fusion protein (Pfu-Sso7d) and the amino acid sequence of the fusion protein are shown in SEQ ID NOs: 7 and 8, respectively.

Construction of Sac7d-ΔTaq fusion.

[0080] A fourth fusion protein was constructed, which joined a sequence-non-specific DNA binding protein from a different species to ΔTaq. Two primers were used to PCR amplify the Sac7d gene from genomic DNA of *Sulfolobus acidocaldarius*. The primers introduced a unique EcoRI site and a unique SpeI site to the PCR fragment at the 5' and 3' termini, respectively. Upon restriction digestion with EcoRI and SpeI, the PCR fragment was ligated into pYW1 (described above) at the corresponding sites. The resulting plasmid expresses a single polypeptide containing the Sac7d protein fused to the N terminus of ΔTaq via the same linker as used in Sso7d-ΔTaq. The DNA sequence of the fusion protein (Sac7d-ΔTaq) and the amino acid sequence of the protein are shown in SEQ ID. NOs: 9 and 10, respectively.

Construction of PL-ΔTaq fusion.

[0081] A fifth fusion protein joins a peptide composed of 14 lysines and 2 arginines to the N terminus of ΔTaq. To generate the polylysine (PL)-ΔTaq fusion protein, two 67 nt oligonucleotides were annealed to form a duplexed DNA fragment with a 5' protruding end compatible with an EcoRI site, and a 3' protruding end compatible with an SpeI site. The DNA fragment encodes a lysine-rich peptide of the following composition: NSKKKKKKKRKKRKKKGGGVT. The numbers of lysines and arginines in this peptide are identical to those in Sso7d. This DNA fragment was ligated into pYW1, predigested with EcoRI and SpeI, to replace the region encoding Sso7d. The resulting plasmid (pLST) expresses a single polypeptide containing the lysine-rich peptide fused to the N terminus of ΔTaq. The DNA sequence encoding the fusion protein (PL-ΔTaq) and the amino acid sequence of the protein are shown in SEQ ID NOs: 11 and NO. 12, respectively.

Example 2. Assessing the processivity of the fusion polymerases.

[0082] This example illustrates enhancement of processivity of the fusion proteins of the invention generated in Example 1.

Polymerase unit definition assay

[0083] The following assay was used to define a polymerase unit. An oligonucleotide was pre-annealed to ssM13mp18 DNA in the presence of Mg^{++}-free reaction buffer and dNTPs. The DNA polymerase of interest was added to the primed DNA mixture. $MgCl_2$ was added to initiate DNA synthesis at 72°C. Samples were taken at various time points and added to TE buffer containing PicoGreen (Molecular Probes, Eugene Oregon). The amount of DNA synthesized was quantified using a fluorescence plate reader. The unit activity of the DNA polymerase of interest was determined by comparing its initial rate with that of a control DNA polymerase (*e.g.*, a commercial polymerase of known unit concentration).

Processivity assay

[0084] Processivity was measured by determining the number of nucleotides incorporated during a single binding event of the polymerase to a primed template.

[0085] Briefly, 40 nM of a 5' FAM-labeled primer (34 nt long) was annealed to 80 nM of circular or linearized ssM13mp18 DNA to form the primed template. The primed template was mixed with the DNA polymerase of interest at a molar ratio of approximately 4000:1 (primed DNA:DNA polymerase) in the presence of standard PCR buffer (free of Mg^{++})and 200 μM of each dNTPs. $MgCl_2$ was added to a final concentration of 2 mM to initiate DNA

synthesis. At various times after initiation, samples were quenched with sequencing loading dye containing 99% formamide, and analyzed on a sequencing gel. The median product length, which is defined as the product length above or below which there are equal amounts of products, was determined based on integration of all detectable product peaks. At a polymerase concentration for which the median product length change with time or polymerase concentration, the length corresponds to the processivity of the enzyme. The ranges presented in Table 1 represent the range of values obtained in several repeats of the assay.

Table 1. Comparison of processivity

DNA polymerase	Median product length (nt)
ΔTaq	2-6
Sso7d-ΔTaq	39-58
PL-ΔTaq	2-6
Taq	15-20
Sso7d-Taq	130-160
Pfu	2-3
Pfu-Sso7d	35-39

[0086] In comparing the processivity of modified enzyme to the unmodified enzyme, ΔTaq had a processivity of 2-6 nucleotides, whereas Sso7d-ΔTaq fusion exhibited a processivity of 39-58 nucleotides (Table I). Full length Taq had a processivity of 15-20 nucleotides, which was significantly lower than that of Sso7d-Taq fusion with a processivity of 130-160 nucleotides. These results demonstrate that Sso7d joined to Taq polymerase enhanced the processivity of the polymerase.

[0087] Pfu belongs to family B of polymerases. Unlike Taq polymerase, Pfu possesses a 3' to 5' exonuclease activity, allowing it to maintain high fidelity during DNA synthesis. A modified Pfu polymerase, in which Sso7d is fused to the C terminus of the full length Pfu polymerase, and an unmodified Pfu polymerase were analyzed in the processivity assay described above. As shown in Table I, the Pfu polymerase exhibited a processivity of 2-3 nt, whereas the Pfu-Sso7d fusion protein had a processivity of 35-39 nt. Thus, the fusion of Sso7d to the C terminus of Pfu resulted in a >10-fold enhancement of the processivity over the unmodified enzyme.

Example 3. Effect of fusion proteins on oligonucleotide annealing temperature

[0088] This experiment demonstrates the increased efficiency of the Sso7d-ΔTaq fusion protein, compared to Taq, to produce product at higher annealing temperatures by stabilizing dsDNA.

[0089] Two primers, primer 1008 (19mer; T_M =56.4°C) and 2180R (20mer; T_M =56.9°C), were used to amplify a 1 kb fragment (1008-2180) of the Taq pol gene. A gradient thermal cycler (MJ Research, Waltham MA) was used to vary the annealing temperature from 50°C to 72°C in a PCR cycling program. The amounts of PCR products generated using identical number of units of Sso7d-ΔTaq and Taq were quantified and compared. The results are shown in Table 2. The Sso7d-ΔTaq fusion protein exhibited significantly higher efficiency than full length Taq at higher annealing temperatures. Thus, the presence of Sso7d in *cis* increases the melting temperature of the primer on the template.

[0090] The annealing temperature assay above was used to investigate whether PL-ΔTaq has any effect on the annealing temperature of primer during PCR amplification. As shown in Table 2 little or no amplified product was observed when the annealing temperature was at or above 63°C.

Table 2. Comparison of activities at different annealing temperatures.

Polymerase	Activity at 63°C	Activity at 66°C	Activity at 69°C
Taq	85%	30%	<10%
Sso7d-ΔTaq	>95%	70%	40%
PL-ΔTaq	<5%	nd	nd

nd: not detectable.

Example 4. Effect of fusion proteins on required primer length

[0091] An enhancement of T_M of the primers (as shown above) predicts that shorter primers could be used by Sso7d-ΔTaq, but not by Taq, to achieve efficient PCR amplification. This analysis shows that Sso7d-ΔTaq is more efficient in an assay using shorter primers compared to Taq.

[0092] Primers of different lengths were used to compare the efficiencies of PCR amplification by Sso7d-ΔTaq and by Taq. The results are shown in Table 3. When two long primers, 57F (22mer, T_M =58°C) and 732R (24mer, T_M =57°C) were used, no significant difference was observed between Sso7d-ΔTaq and Taq at either low or high annealing temperatures. When medium length primers, 57F15 (15mer, T_M =35°C) and 732R16 (16mer, T_m=35°C), were used, Sso7d-ΔTaq was more efficient than Taq, especially when the

annealing temperature was high. The most striking difference between the two enzymes was observed with short primers, 57F12 (12mer) and 732R16 (16mer), where Sso7d-ΔTaq generated 10 times more products than Taq at both low and high annealing temperatures.

[0093] PCR using primers 57F12 (12 nt) and 732R16 (16 nt) were used to compare the efficiency of Sac7d-ΔTaq to the unmodified full length Taq in PCR reaction. Similar to Sso7d-ΔTaq, Sac7d-ΔTaq is significantly more efficient than Taq in amplifying using short primers.

[0094] A primer length assay was used to determine the ability of PL-ΔTaq to use short primers in PCR amplification. When long primers (57F and 732R) were used, the amplified product generated by PL-ΔTaq is ~50% of that by Sso7d-ΔTaq. When short primers (57F12 and 732R16) were used, the amplified product generated by PL-ΔTaq is <20% of that by Sso7d-ΔTaq .

Table 3. Comparison of the effect of primer length on PCR amplification by Sso7d-ΔTaq and Taq DNA polymerase.

polymerase	22 nt primer		15 nt primer		12 nt primer	
	Anneal @55°C	Anneal @63°C	Anneal @49°C	Anneal @54°C	Anneal @49°C	Anneal @54°C
Taq	14000	9000	5500	<500	1000	undetectable
Sso7d-ΔTaq	17000	13000	15000	5000	10000	3000
Sso7d-ΔTaq:Taq	1.2:1	1.4:1	2.7:1	>10:1	10:1	>10:1

Increased performance of fusion polymerases in PCR reactions

[0095] The increased stability and/or processivity of the fusion proteins of the invention provide increased efficiency in performing various modification reactions. For example, polymerase fusion proteins can provide more efficient amplification in PCR reactions. Many factors influence the outcome of a PCR reaction, including primer specificity, efficiency of the polymerase, quality, quantity and GC-content of the template, length of the amplicon, *etc.* Examples 5-8 demonstrate that fusion proteins that include a double-stranded sequence-non-specific nucleic acid binding domain, *e.g.*, Sso7d, joined to a thermostable polymerase or polymerase domain have several advantageous features over the unmodified enzyme in PCR applications.

Example 5. Sso7d fusion proteins exhibit a higher and broader salt-tolerance in PCR

[0096] The binding of polymerase to a primed DNA template is sensitive to the ionic strength of the reaction buffer due to electrostatic interactions, which is stronger in low salt concentration and weaker in high. The presence of Sso7d in a fusion polymerase protein stabilizes the binding interaction of the polymerase to DNA template. This example demonstrates that Sso7d fusion proteins exhibit improved performance in PCR reactions containing elevated KCl concentrations.

[0097] Lambda DNA (2 pM) was used as a template in a PCR reactions with primers 57F and 732R. The concentration of KCl was varied from 10 mM to 150 mM, while all other components of the reaction buffer were unchanged. The PCR reaction was carried out using a cycling program of 94°C for 3 min, 20 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec, followed by 72°C for 10 min. Upon completion of the reaction, 5 μl of the PCR reaction was removed and mixed with 195 μl of 1:400 dilution of PicoGreen in TE to quantify the amounts of amplicon generated. The PCR reaction products were also analyzed in parallel on an agarose gel to verify that amplicons of expected length were generated (data not shown). The effects of KCl concentration on the PCR efficiency of Sso7d-ΔTaq versus that of ΔTaq, and Pfu-Sso7d versus Pfu are shown in Table 4. The unmodified enzymes, ΔTaq and Pfu, showed a preference for KCl concentration below 25 mM and 40 mM, respectively, to maintain 80% of the maximum activity. In contrast, fusion proteins Sso7d-ΔTaq and Pfu-Sso7d maintain 80% of the maximum activity in 30-100 mM and 60-100 mM KCl, respectively. Thus, the Sso7d fusion proteins were more tolerant of elevated KCl concentration in comparison to their unmodified counter parts. This feature of the hybrid polymerase will potentially allow PCR amplification from low quality of DNA template, *e.g.*, DNA samples prepared from, but not limited to, blood, food, and plant sources.

Table 4. Sso7d modification increases salt-tolerance of polymerase in PCR reaction

Enzyme	**Enzyme concentration**	**[KCl] for 80% activity**
ΔTaq	20U/ml	<25 mM
Sso7d-ΔTaq	20U/ml	30-100 mM
Pfu	3 U/ml	<40 mM
Pfu-Sso7d	12U/ml* (equal molar)	60-100 mM

* Pfu-Sso7d has a 4-fold higher specific activity than Pfu. The specific activity is defined as unit/mol of enzyme.

Example 6. Sso7d-fusion polymerases are more tolerant to SYBR Green I in real-time PCR

[0098] Three pairs of unmodified and modified enzymes were compared: commercial ΔTaq (ABI, Foster City, CA) vs. Sso7d-ΔTaq, Taq vs. Sso7d-Taq, and commercial Pfu (Stratagene, La Jolla CA) vs. Pfu-Sso7d. In addition to the 20U/ml concentration used for all enzymes, a 5-fold higher concentration (100 U/ml) of ΔTaq and Pfu were used as well. The Ct values represent the number of cycles required to generate a detectable amount of DNA, and thus a lower Ct value reflects a greater amplification efficiency for the enzyme. Consistent Ct values are also preferable, indicating the reaction is robust to differences in dye concentration. Two extension times (10s and 30s) were used. The SYBR Green 1 concentration is indicated as 0.5x, etc. The 1x SYBR Green I is defined as a SYBR Green I solution in TE (10mM Tris pH 7.5, 1mM EDTA) that has an absorbance at 495 nm of 0.40±0.02. SYBR Green I was purchased from Molecular Probes (Eugene, Oregon) as a 10,000x stock in DMSO. In all three pairs, the modified polymerase showed significantly higher tolerance of dye. The differences are most striking in the case of ΔTaq vs. Sso7d-ΔTaq.

Table 5. Sso7d fusion proteins are more tolerant of SYBR Green I. (The symbol "--" indicates that no amplification was observed in 40 cycles

	10s @ 72°C										
	MgCl2	**2 mM**					**3 mM**				
		SYBR Green I					**SYBR Green I**				
ENZYMES	**Unit/ml**	0.5x	1x	1.5x	2x	2.5x	0.5x	1x	1.5x	2x	2.5x
ΔTaq	20	--	--	--	--	--	--	--	--	--	--
ΔTaq	100	--	--	--	--	--	--	--	--	--	--
Sso7d-ΔTaq	20	23.3	22.5	22.5	22.3	22.4	22.9	22.2	22	22.2	21.8
Taq	20	23	23.6	--	--	--	22.5	22.3	22.6	--	--
Sso7d-Taq	20	23.3	23.3	23.2	23.5	--	24	24	23.1	23.4	23.6
Pfu	20	31.2	--	--	--	--	31.5	--	--	--	--
Pfu	100	21.8	25	--	--	--	22.6	23.3	30	--	--
Pfu-Sso7d	20	21.5	22.3	35	--	--	21.8	22	22.6	27.2	--
	30s @ 72°C										
ΔTaq	20	--	--	--	--	--	--	--	--	--	--
ΔTaq	100	--	--	--	--	--	26.8	--	--	--	--
Sso7d-ΔTaq	20	23.8	22.3	22.6	21.8	21.7	22.3	21	21.3	21.8	21.8
Taq	20	24.2	24.6	29.4	∞	--	22.8	22.1	22.6	25	--
Sso7d-Taq	20	24.2	23.5	23	22.7	24.2	24.7	23.1	23.6	23.1	22.9
Pfu	20	33.2	--	--	--	--	29.4	--	--	--	--
Pfu	100	27.6	30.6	--	--	--	24.8	29.8	--	--	--
Pfu-Sso7d	20	25	24.8	25.4	24.4	--	23.1	25.3	23.6	26.1	--

Example 7. Sso7d-fusion polymerases are more tolerant to crude template preparations

A. Resistance to bacterial contamination in PCR

[0099] Colony PCR is a useful technique in which small samples of single bacterial colonies are lysed and added directly to PCR reactions for the purpose of screening the colonies for particular DNA sequences. Colony PCR has a high failure rate, presumably because of contaminants carries over from the colony. Polymerases resistant to cell extracts are desirable because they presumably will be more successful in colony PCR.

Materials For "Dirty "PCR
Lambda template (10 ng/ml): amplicon is a 891 bp fragment
Primers 56F/55R (T_M 56° and 55°),400 nM
Enzymes: Sst (Sso7d-ΔTaq) vs. PE Stf (ΔTaq), STq (Sso7d-Taq) vs. Taq-HIS or AmpliTaq or Amersham Taq, and Stratagene Pfu vs. PfS (Pfu-Sso7d) All enzymes are 20 U/ml except where indicated
200 μM each dNTP
2 mM $MgCl_2$, except 1.5 mM for Amersham Taq and AmpliTaq
Reactions were 20 μl

Methods:

[0100] *E. coli* were grown to saturation, spun down, suspended in water at an OD of 100, and frozen and thawed to disrupt cells. Dilutions of the disrupted bacteria were added at various concentrations to PCR reactions containing lambda DNA as template and two primers to amplify a 890 bp amplicon. 1X is equivalent to an OD of 10 (10 OD units/ml). The cycling conditions were as follows:

1) 95°C -20"
2) 94°C -5"
3) 60°C -15"
4) 72°C -45"
5) repeat steps 2-4 19 times
6) 72°C -5'
7) 4°C forever
8) END

[0101] The experiment showed that Sso7d-ΔTaq significantly out performed Stoffel fragment (Applied Biosystems, Foster City, CA). Stoffel (Stf) is a trade name for a preparation of ΔTaq. Using 20 U/ml enzyme in the final reaction, Sso7d-ΔTaq allowed PCR amplification in the presence of 0.25x of cell dilution. When the same unit concentrations of Stoffel was used, no detectable product was generated, even in the most dilute cell solution. When 220 u/ml Stoffel was used, a detectable amount of product was generated at a 0.06x or

lower concentration of the cell dilution. Thus, the resistance of Sso7d-ΔTaq to bacterial contamination in PCR reaction is more than 10-folder higher than that of the unmodified enzyme Stoffel.

[0102] Similarly, Pfu-Sso7d showed more resistance to bacterial contamination than Pfu, although both enzymes appeared to be more sensitive to the contamination than Taq-based enzymes. With 20 U/ml enzyme in the final reaction, Pfu allowed amplification only in the presence of 0.00006x or lower concentrations of cell dilution. In contrast, Pfu-Sso7d allowed efficient PCR amplification in 0.002x of cell dilution. Thus, Pfu-Sso7d has a 30-fold higher tolerance to bacterial contamination in PCR than the unmodified enzyme Pfu.

B. Resistance to plant and blood contamination in PCR

[0103] The same problems exist with other crude template preparations. PCR fails due to contaminants carried over in the template preparation. This example shows results with crude plant and blood preps. Dilution series were made of plant leaf homogenate from *Fritallaria agrestis*, a species of lily, and whole human blood. Dilutions were made with 1 x TE, pH 8.0 at 1/10, 1/100, 1/1000. One microliter of a dilution was added to the appropriate reaction mix. The PCR cycling protocol was as follows:

94°C 2 min
94°C 10 sec
59°C 20 sec for Taq & Sso7d-Taq (54°C for Pfu & Pfu-Sso7d)
72°C 30 sec
repeat cycle 34 times
72°C 10 min

[0104] The reaction products were analyzed on agarose gels (Figure 1A and Figure 1B). Figure 1A shows a comparison of the contamination resistance of Pfu vs. PfS. Lanes 1-4 and 14-17 show progressive 10-fold dilutions of plant leaf homogenate. Pfu shows significant inhibition by a 1:10 dilution (lane 2), while PfS is completely resistant to this dilution (lane 7). Similarly, lanes 6-9 and 19-22 show progressive 10-fold dilutions of blood. Pfu is significantly inhibited by 1 microliter of blood, while PfS is resistant. Lanes 10 and 23 are positive controls (no plant or blood), while lanes 11 and 24 are negative controls (no plant or blood or template).

[0105] Figure 1B shows a comparison between Taq and Sso7d-Taq. The upper panel shows reactions performed with 20U/ml Taq, and the lower panel shows reactions performed with 20U/ml Sso7d-Taq. Lanes 1-4 in each panel show progressive 10-fold dilutions of plant leaf homogenate and lanes 7-10 show progressive 10-fold dilutions of blood. Sso7d-Taq can

amplify a product even in the presence of 1 μl whole blood, while Taq is inhibited by 100-fold less blood. Lanes 5 are positive controls (no plant or blood), while lanes 11 are negative controls (no plant or blood or template).

Example 8. Sso7d-fusion polymerases have advantages in cycle sequencing

[0106] Plasmid clones encoding improved polymerases suitable for DNA sequencing have been constructed, and the protein products have been purified. and purified. The first enzyme is Sso7d-ΔTaq(Y), (SEQ ID No: 30 and 31 with mutations indicated in bold font) which is the same as the enzyme Sso7d-ΔTaq, except modified according to the method of Tabor and Richardson (US Patent No. 5,614,365) to have a "Y" substituted for an "F" residue at the indicated position in SEQ ID NO:31. The second enzyme is Sso7d-ΔTaq(E5;Y) (SEQ ID No: 32 and 33) with mutations indicated in bold font) which is the same as Sso7d-Taq, except modified according to the method of Tabor and Richardson and also containing point mutations that inactivate the 5'-3' nuclease domain.

[0107] The processivity of each Sso7d fusion polymerase was compared to its unmodified counterpart, *i.e.*, the polymerase without the Sso7d domain. The results in Table 6 show that the Sso7d fusion polymerases are more processive.

Table 6.

Median Processivity Product Length at 10 mM KCl	
ΔTaq (Y)	3 to 4 nts.
Sso7d-ΔTaq (Y)	11 to 13 nts.
ΔTaq (E5)(Y)	5 to 6 nts.
Sso7d-ΔTaq (E5)(Y)	34 to 47 nts.

[0108] Sequencing reactions using the fusion polymerases and their unmodified counterparts were performed by separating the components of a commercial sequencing kit (BigDye terminator Kit v.3, ABI, Foster City CA). Low-molecular-weight components were separated from the enzymes by ultrafiltration. Sequencing reactions performed by combining the low-molecular-weight fraction with the improved enzymes showed good signal strength vs. base number curves. Furthermore, the improved polymerases, *e.g.*, Sso7d-ΔTaq(E5;Y), was able to continued through a hard stop better the other enzymes. Such an improved

polymerase is also able to continue through dinucleotide, trinucletide, and long single base repeats more effectively than a counterpart polymerase..

[0109] Optimization of the sequencing reactions will demonstrate improvements in peak height evenness, contamination resistance, and lowered requirement for template and/or enzyme concentration.

Table of sequences

SEQ ID NO:1 Synthetic Sso7d gene

GCAACCGTAAAGTTCAAGTACAAAGGCGAAGAAAAAGAGGTAGACATCTCCAA GATCAAGAAAGTATGGCGTGTGGGCAAGATGATCTCCTTCACCTACGACGAGGG CGGTGGCAAGACCGGCCGTGGTGCGGTAAGCGAAAAGGACGCGCCGAAGGAGC TGCTGCAGATGCTGGAGAAGCAGAAAAAG

SEQ ID NO:2 The amino acid sequence of Sso7d

ATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQ MLEKQKK

SEQ ID NO:3 The DNA sequence encoding the Sso7d-ΔTaq fusion protein

ATGATTACGAATTCGAGCGCAACCGTAAAGTTCAAGTACAAAGGCGAAGAAAAA GAGGTAGACATCTCCAAGATCAAGAAAGTATGGCGTGTGGGCAAGATGATCTCC TTCACCTACGACGAGGGCGGTGGCAAGACCGGCCGTGGTGCGGTAAGCGAAAAG GACGCGCCGAAGGAGCTGCTGCAGATGCTGGAGAAGCAGAAAAAGGGCGGCGG TGTCACTAGTCCCAAGGCcCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCC TTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCT GGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCT CAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGC CCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTAC CTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGG GAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCC AACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAG GTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGC CTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCC GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCG GGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAA GACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCG CGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCT GAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCG CCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTC CGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGC

CGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAG
ATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCT
TCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCC
CCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCG
GGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTA
CGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGG
GCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGAC
CCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGT
GCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGC
CGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGG
GGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGA
GAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTATCC
CCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGC
CAAGGAGGGCATTGATGGCCGCGGCGGAGGCGGGCATCATCATCATCATCATTA
A

SEQ ID NO:4 The amino acid sequence of Sso7d-ΔTaq fusion protein

MITNSSATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDA
PKELLQMLEKQKKGGGVTSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAA
ARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDP
SNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLS
AVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLF
DELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIH
PRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYS
QIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGV
LYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
RYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLL
QVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKEGIDGR
GGGGHHHHHH

SEQ ID NO:5 The DNA sequence encoding the Sso7d-Taq fusion protein

ATGATTACGAATTCGAGCGCAACCGTAAAGTTCAAGTACAAAGGCGAAGAAAAA
GAGGTAGACATCTCCAAGATCAAGAAAGTATGGCGTGTGGGCAAGATGATCTCC

TTCACCTACGACGAGGGCGGTGGCAAGACCGGCCGTGGTGCGGTAAGCGAAAAG
GACGCGCCGAAGGAGCTGCTGCAGATGCTGGAGAAGCAGAAAAAGGGCGGCGG
TGTCACTAGTGGGATGCTGCCCCTCTTTGAGCCCAAGGGCCGGGTCCTCCTGGTG
GACGGCCACCACCTGGCCTACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCA
GCCGGGGGGAGCCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTGACGCCAAGGCCCCCT
CCTTCCGCCACGAGGCCTACGGGGGGTACAAGGCGGGCCGGGCCCCCACGCCAG
AGGACTTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCT
GGCGCGCCTCGAGGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCAGCCTGGC
CAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAAG
ACCTTTACCAGCTCCTTTCCGACCGCATCCACGTCCTCCACCCCGAGGGGTACCT
CATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAGTGGGC
CGACTACCGGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGG
CATCGGGGAGAAGACGGCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAG
CCCTCCTCAAGAACCTGGACCGGCTGAAGCCCGCCATCCGGGAGAAGATCCTGG
CCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACCT
GCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAG
GGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTG
GAAAGCCCCAAGGCcCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTC
GTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGG
CCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCA
GGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCC
TGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCT
CCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGA
GTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAA
CCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGT
GGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCT
GGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCG
CCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGG
GACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAG
ACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGC
GAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTG
AAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCC
TCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCG

ATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCG
GGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGAT
AGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTC
CAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCC
CGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTTCGGG
GTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACG
AGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGC
CTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCC
TCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGC
GGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCG
ACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGG
CCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGA
GGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTATCCCC
TGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCA
AGGAGGGCATTGATGGCCGCGGCGGAGGCGGGCATCATCATCATCATCATTAA

SEQ ID NO:6 The amino acid sequence of Sso7d-Taq fusion protein

MITNSSATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDA
PKELLQMLEKQKKGGGVTSGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGE
PVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFRHEAYGGYKAGRAPTPEDFPRQ
LALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLSDR
IHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLL
EEWGSLEALLKNLDRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREP
DRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADL
LALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREV
ERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQL
ERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPL
PDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVA
LDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTI
NFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETL
FGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGA
RMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKE
GIDGRGGGGHHHHHH

SEQ ID NO:7 The DNA sequence encoding the Pfu-Sso7d fusion protein

ATGATTTTAGATGTGGATTACATAACTGAAGAAGGAAAACCTGTTATTAGGCTAT
TCAAAAAAGAGAACGGAAAATTTAAGATAGAGCATGATAGAACTTTTAGACCAT
ACATTTACGCTCTTCTCAGGGATGATTCAAAGATTGAAGAAGTTAAGAAAATAAC
GGGGGAAAGGCATGGAAAGATTGTGAGAATTGTTGATGTAGAGAAGGTTGAGAA
AAAGTTTCTCGGCAAGCCTATTACCGTGTGGAAACTTTATTTGGAACATCCCCAA
GATGTTCCCACTATTAGAGAAAAAGTTAGAGAACATCCAGCAGTTGTGGACATCT
TCGAATACGATATTCCATTTGCAAAGAGATACCTCATCGACAAAGGCCTAATACC
AATGGAGGGGGAAGAAGAGCTAAAGATTCTTGCCTTCGATATAGAAACCCTCTA
TCACGAAGGAGAAGAGTTTGGAAAAGGCCCAATTATAATGATTAGTTATGCAGA
TGAAAATGAAGCAAAGGTGATTACTTGGAAAAACATAGATCTTCCATACGTTGA
GGTTGTATCAAGCGAGAGAGAGATGATAAAGAGATTTCTCAGGATTATCAGGGA
GAAGGATCCTGACATTATAGTTACTTATAATGGAGACTCATTCGACTTCCCATAT
TTAGCGAAAAGGGCAGAAAAACTTGGGATTAAATTAACCATTGGAAGAGATGGA
AGCGAGCCCAAGATGCAGAGAATAGGCGATATGACGGCTGTAGAAGTCAAGGG
AAGAATACATTTCGACTTGTATCATGTAATAACAAGGACAATAAATCTCCCAACA
TACACACTAGAGGCTGTATATGAAGCAATTTTTGGAAAGCCAAAGGAGAAGGTA
TACGCCGACGAGATAGCAAAAGCCTGGGAAAGTGGAGAGAACCTTGAGAGAGTT
GCCAAATACTCGATGGAAGATGCAAAGGCAACTTATGAACTCGGGAAAGAATTC
CTTCCAATGGAAATTCAGCTTTCAAGATTAGTTGGACAACCTTTATGGGATGTTT
CAAGGTCAAGCACAGGGAACCTTGTAGAGTGGTTCTTACTTAGGAAAGCCTACG
AAAGAAACGAAGTAGCTCCAAACAAGCCAAGTGAAGAGGAGTATCAAAGAAGG
CTCAGGGAGAGCTACACAGGTGGATTCGTTAAAGAGCCAGAAAAGGGGTTGTGG
GAAAACATAGTATACCTAGATTTTAGAGCCCTATATCCCTCGATTATAATTACCC
ACAATGTTTCTCCCGATACTCTAAATCTTGAGGGATGCAAGAACTATGATATCGC
TCCTCAAGTAGGCCACAAGTTCTGCAAGGACATCCCTGGTTTTATACCAAGTCTC
TTGGGACATTTGTTAGAGGAAAGACAAAAGATTAAGACAAAAATGAAGGAAACT
CAAGATCCTATAGAAAAAATACTCCTTGACTATAGACAAAAAGCGATAAAACTC
TTAGCAAATTCTTTCTACGGATATTATGGCTATGCAAAAGCAAGATGGTACTGTA
AGGAGTGTGCTGAGAGCGTTACTGCCTGGGGAAGAAAGTACATCGAGTTAGTAT
GGAAGGAGCTCGAAGAAAAGTTTGGATTTAAAGTCCTCTACATTGACACTGATG
GTCTCTATGCAACTATCCCAGGAGGAGAAAGTGAGGAAATAAAGAAAAAGGCTC
TAGAATTTGTAAAATACATAAATTCAAAGCTCCCTGGACTGCTAGAGCTTGAATA

TGAAGGGTTTTATAAGAGGGGATTCTTCGTTACGAAGAAGAGGTATGCAGTAATAGATGAAGAAGGAAAAGTCATTACTCGTGGTTTAGAGATAGTTAGGAGAGATTGGAGTGAAATTGCAAAAGAAACTCAAGCTAGAGTTTTGGAGACAATACTAAAACACGGAGATGTTGAAGAAGCTGTGAGAATAGTAAAAGAAGTAATACAAAAGCTTGCCAATTATGAAATTCCACCAGAGAAGCTCGCAATATATGAGCAGATAACAAGACCATTACATGAGTATAAGGCGATAGGTCCTCACGTAGCTGTTGCAAAGAAACTAGCTGCTAAAGGAGTTAAAATAAAGCCAGGAATGGTAATTGGATACATAGTACTTAGAGGCGATGGTCCAATTAGCAATAGGGCAATTCTAGCTGAGGAATACGATCCCAAAAAGCACAAGTATGACGCAGAATATTACATTGAGAACCAGGTTCTTCCAGCGGTACTTAGGATATTGGAGGGATTTGGATACAGAAAGGAAGACCTCAGATACCAAAAGACAAGACAAGTCGGCCTAACTTCCTGGCTTAACATTAAAAAATCCGGTACCGGCGGTGGCGGTGCAACCGTAAAGTTCAAGTACAAAGGCGAAGAAAAAGAGGTAGACATCTCCAAGATCAAGAAAGTATGGCGTGTGGGCAAGATGATCTCCTTCACCTACGACGAGGGCGGTGGCAAGACCGGCCGTGGTGCGGTAAGCGAAAAGGACGCGCCGAAGGAGCTGCTGCAGATGCTGGAGAAGCAGAAAAAGTGA

SEQ ID NO:8 The amino acid sequence of the Pfu-Sso7d fusion protein

MILDVDYITEEGKPVIRLFKKENGKFKIEHDRTFRPYIYALLRDDSKIEEVKKITGERHGKIVRIVDVEKVEKKFLGKPITVWKLYLEHPQDVPTIREKVREHPAVVDIFEYDIPFAKRYLIDKGLIPMEGEEELKILAFDIETLYHEGEEFGKGPIIMISYADENEAKVITWKNIDLPYVEVVSSEREMIKRFLRIIREKDPDIIVTYNGDSFDFPYLAKRAEKLGIKLTIGRDGSEPKMQRIGDMTAVEVKGRIHFDLYHVITRTINLPTYTLEAVYEAIFGKPKEKVYADEIAKAWESGENLERVAKYSMEDAKATYELGKEFLPMEIQLSRLVGQPLWDVSRSSTGNLVEWFLLRKAYERNEVAPNKPSEEEYQRRLRESYTGGFVKEPEKGLWENIVYLDFRALYPSIIITHNVSPDTLNLEGCKNYDIAPQVGHKFCKDIPGFIPSLLGHLLEERQKIKTKMKETQDPIEKILLDYRQKAIKLLANSFYGYYGYAKARWYCKECAESVTAWGRKYIELVWKELEEKFGFKVLYIDTDGLYATIPGGESEEIKKKALEFVKYINSKLPGLLELEYEGFYKRGFFVTKKRYAVIDEEGKVITRGLEIVRRDWSEIAKETQARVLETILKHGDVEEAVRIVKEVIQKLANYEIPPEKLAIYEQITRPLHEYKAIGPHVAVAKKLAAKGVKIKPGMVIGYIVLRGDGPISNRAILAEEYDPKKHKYDAEYYIENQVLPAVLRILEGFGYRKEDLRYQKTRQVGLTSWLNIKKSGTGGGGATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDAPKELLQMLEKQKK

SEQ ID NO:9 The DNA sequence encoding the Sac7d-ΔTaq fusion protein

ATGATTACGAATTCGACGGTGAAGGTAAAGTTCAAGTATAAGGGTGAAGAGAAA
GAAGTAGACACTTCAAAGATAAAGAAGGTTTGGAGAGTAGGCAAAATGGTGTCC
TTTACCTATGACGACAATGGTAAGACAGGTAGAGGAGCTGTAAGCGAGAAAGAT
GCTCCAAAAGAATTATTAGACATGTTAGCAAGAGCAGAAAGAGAGAAGAAAGG
CGGCGGTGTCACTAGTCCCAAGGCCCTGGAGGAGGCCCCCTGGCCCCCGCCGGA
AGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTT
CTGGCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTAT
AAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGC
GTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCC
TCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTA
CGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGC
TCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTA
CCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGG
GGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGA
GATCGCCCGCCTCGAGGCCGGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCA
ACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCAT
CGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGC
CCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCAC
CAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACG
GGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGT
AGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGA
TCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAG
CCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCG
GGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGG
CGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAA
CTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATC
CCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGG
TGCGGGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTG
GAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAG
AGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACC
GCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAA
ATGGGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCA

AAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGT
GTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCT
CTCCGCCAAGGAGGGCATTGATGGCCGCGGCGGAGGCGGGCATCATCATCATCA
TCATTAA

SEQ ID NO:10 The amino acid sequence of the Sac7d-ΔTaq fusion protein

MITNSTVKVKFKYKGEEKEVDTSKIKKVWRVGKMVSFTYDDNGKTGRGAVSEKDA
PKELLDMLARAEREKKGGGVTSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLAL
AAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLL
DPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERP
LSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERV
LFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLI
HPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDY
SQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFG
VLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGR
RRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARML
LQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKEGIDG
RGGGHHHHHH

SEQ ID NO:11 The DNA sequence encoding the PL-ΔTaq fusion protein

ATGATTACGAATTCGAAGAAAAAGAAAAAGAAAAAGCGTAAGAAACGCAAAAA
GAAAAAGAAAGGCGGCGGTGTCACTAGTGGCGCAACCGTAAAGTTCAAGTACAA
AGGCGAAGAAAAAGAGGTAGACATCTCCAAGATCAAGAAAGTATGGCGTGTGG
GCAAGATGATCTCCTTCACCTACGACGAGGGCGGTGGCAAGACCGGCCGTGGTG
CGGTAAGCGAAAAGGACGCGCCGAAGGAGCTGCTGCAGATGCTGGAGAAGCAG
AAAAAGGGCGGCGGTGTCACCAGTCCCAAGGCCCTGGAGGAGGCCCCCTGGCCC
CCGCCGGAAGGGGCCTTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGG
CCGATCTTCTGGCCCTGGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCG
AGCCTTATAAAGCCCTCAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAG
ACCTGAGCGTTCTGGCCCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCC
CATGCTCCTCGCCTACCTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCC
CGGCGCTACGGCGGGGAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCC
GAGAGGCTCTTCGCCAACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTT

TGGCTTTACCGGGAGGTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGG
CCACGGGGGTGCGCCTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGC
CGAGGAGATCGCCCGCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTC
AACCTCAACTCCCGGGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTC
CCGCCATCGGCAAGACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCC
TGGAGGCCCTCCGCGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGG
AGCTCACCAAGCTGAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCC
CAGGACGGGCCGCCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAG
GCTAAGTAGCTCCGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGG
CAGAGGATCCGCCGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTG
GACTATAGCCAGATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAAC
CTGATCCGGGTCTTCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGG
ATGTTCGGCGTCCCCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAG
ACCATCAACTTCGGGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGC
TAGCCATCCCTTACGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTT
CCCCAAGGTGCGGGCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGG
GGTACGTGGAGACCCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCC
GGGTGAAGAGCGTGCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCC
AGGGCACCGCCGCCGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCT
GGAGGAAATGGGGGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGA
GGCCCCAAAAGAGAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGG
AGGGGGTGTATCCCCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGG
ACTGGCTCTCCGCCAAGGAGGGCATTGATGGCCGCGGCGGAGGCGGGCATCATC
ATCATCATCATTAA

SEQ ID NO:12 The amino acid sequence of PL–ΔTaq fusion protein

MITNSKKKKKKKRKKRKKKKKGGGVTSGATVKFKYKGEEKEVDISKIKKVWRVGK
MISFTYDEGGGKTGRGAVSEKDAPKELLQMLEKQKKGGGVTSPKALEEAPWPPPEG
AFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVL
ALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFAN
LWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLE
AEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIV
EKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTP

LGQRIRRAFIAEEGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMF
GVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVR
AWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAAD
LMKLAMVKLFPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPL
AVPLEVEVGIGEDWLSAKEGIDGRGGGGHHHHHH

SEQ ID NO:13 PRIMER L71F

5'-CCTGCTCTGCCGCTTCACGC-3'

SEQ ID NO:14 PRIMER L71R

5'-GCACAGCGGCTGGCTGAGGA-3'

SEQ ID NO:15 PRIMER L18015F

5'-TGACGGAGGATAACGCCAGCAG-3'

SEQ ID NO:16 PRIMER L23474R

5'-GAAAGACGA TGGGTCGCTAATACGC-3'

SEQ ID NO:17 PRIMER L18015F

5'-TGACGGAGGATAAC GCCAGCAG-3'

SEQ ID NO:18 PRIMER L29930R

5'-GGGGTTGGAGGTCAATGGGTTC-3'

SEQ ID NO:19 PRIMER L30350F

5'-CCTGCTCTGCCGCTTCACGC-3'

SEQ ID NO:20 PRIMER L35121R

5'-CACATGGTACAGCAAGCCTGGC-3'

SEQ ID NO:21 PRIMER L2089F

5'-CCCGTATCTGCTGGGA TACTGGC-3

SEQ ID NO: 22 PRIMER L7112R
5'-CAGCGGTGCTGACTGAATCATGG-3

SEQ ID NO:23 PRIMER L30350F
5'-CCTGCCTGCCGCTTCACGC-3'

SEQ ID NO:24 PRIMER L40547R
5'-CCAATACCCGTTTCA TCGCGGC-3'

SEQ ID NO:25 PRIMER H-Amelo-Y
5'-CCACCTCATCCTGG GCACC-3'

SEQ ID NO:26 PRIMER H-Amelo-YR
5'-GCTTGAGGCCAACCATCAGAGC-3'

SEQ ID NO:27 Human beta-globin primer 536F
5'-GGTTGGCCAATCTACTCCCAGG-3'

SEQ ID NO:28 Human beta-globin primer 536R
5'-GCTCACTCAGTGTGGCAAAG-3'

SEQ ID NO:29 Human beta-globin primer 1408R
5'-GATTAGCAAAAGGGCCTAGCTTGG-3'

SEQ ID NO:30 The DNA sequence encoding the Sso7d-ΔTaq(Y) protein
ATGATTACGAATTCGAGCGCAACCGTAAAGTTCAAGTACAAAGGCGAAGAAAAA
GAGGTAGACATCTCCAAGATCAAGAAAGTATGGCGTGTGGGCAAGATGATCTCC
TTCACCTACGACGAGGGCGGTGGCAAGACCGGCCGTGGTGCGGTAAGCGAAAAG
GACGCGCCGAAGGAGCTGCTGCAGATGCTGGAGAAGCAGAAAAAGGGCGGCGG
TGTCACTAGTCCCAAGGCcCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCC
TTCGTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCT
GGCCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCT
CAGGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGC

CCTGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTAC
CTCCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGG
GAGTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCC
AACCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAG
GTGGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGC
CTGGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCC
GCCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCG
GGACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAA
GACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCG
CGAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCT
GAAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCG
CCTCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTC
CGATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGC
CGGGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAG
ATAGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCT
TCCAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCC
CCCGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAAC**TAC**G
GGGTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTA
CGAGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGG
GCCTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGAC
CCTCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGT
GCGGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGC
CGACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGG
GGCCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGA
GAGGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTATCC
CCTGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGC
CAAGGAGGGCATTGATGGCCGCGGCGGAGGCGGGCATCATCATCATCATCATTA
A

SEQ ID NO:31 The amino acid sequence of Sso7d-ΔTaq(Y) protein

MITNSSATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDA
PKELLQMLEKQKKGGGVTSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAA
ARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDP

SNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLS
AVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLF
DELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIH
PRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVALDYS
QIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINYGV
LYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRR
RYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMGARMLL
QVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAKEGIDGR
GGGGHHHHHH

SEQ ID NO:32 The DNA sequence encoding the Sso7d-ΔTaq (E5)(Y) protein

ATGATTACGAATTCGAGCGCAACCGTAAAGTTCAAGTACAAAGGCGAAGAAAAA
GAGGTAGACATCTCCAAGATCAAGAAAGTATGGCGTGTGGGCAAGATGATCTCC
TTCACCTACGACGAGGGCGGTGGCAAGACCGGCCGTGGTGCGGTAAGCGAAAAG
GACGCGCCGAAGGAGCTGCTGCAGATGCTGGAGAAGCAGAAAAAGGGCGGCGG
TGTCACTAGTGGGATGCTGCCCCTCTTTGAGCCCAAGGGCCGGGTCCTCCTGGTG
GACGGCCACCACCTGGCCTACCGCACCTTCCACGCCCTGAAGGGCCTCACCACCA
GCCGGGGGGAGCCGGTGCAGGCGGTCTACGGCTTCGCCAAGAGCCTCCTCAAGG
CCCTCAAGGAGGACGGGGACGCGGTGATCGTGGTCTTTGACGCCAAGGCCCCCT
CCTTCCCCCACGAGGCCTACGGGGGGCACAAGGCGGGCCGGGCCCCCACGCCAG
AGGACTTTCCCCGGCAACTCGCCCTCATCAAGGAGCTGGTGGACCTCCTGGGGCT
GGCGCGCCTCGAGGTCCCGGGCTACGAGGCGGACGACGTCCTGGCCAGCCTGGC
CAAGAAGGCGGAAAAGGAGGGCTACGAGGTCCGCATCCTCACCGCCGACAAAG
ACCTTTACCAGCTCCTTTCCGACCGCATCCACGTCCTCCACCCCGAGGGGTACCT
CATCACCCCGGCCTGGCTTTGGGAAAAGTACGGCCTGAGGCCCGACCAGTGGGC
CGACTACCGGGCCCTGACCGGGGACGAGTCCGACAACCTTCCCGGGGTCAAGGG
CATCGGGGAGAAGACGGCGAGGAAGCTTCTGGAGGAGTGGGGGAGCCTGGAAG
CCCTCCTCAAGAACCTGGACCGGCTGAAGCCCGCCATCCGGGAGAAGATCCTGG
CCCACATGGACGATCTGAAGCTCTCCTGGGACCTGGCCAAGGTGCGCACCGACCT
GCCCCTGGAGGTGGACTTCGCCAAAAGGCGGGAGCCCGACCGGGAGAGGCTTAG
GGCCTTTCTGGAGAGGCTTGAGTTTGGCAGCCTCCTCCACGAGTTCGGCCTTCTG
GAAAGCCCCAAGGCcCTGGAGGAGGCCCCCTGGCCCCCGCCGGAAGGGGCCTTC
GTGGGCTTTGTGCTTTCCCGCAAGGAGCCCATGTGGGCCGATCTTCTGGCCCTGG

CCGCCGCCAGGGGGGGCCGGGTCCACCGGGCCCCCGAGCCTTATAAAGCCCTCA
GGGACCTGAAGGAGGCGCGGGGGCTTCTCGCCAAAGACCTGAGCGTTCTGGCCC
TGAGGGAAGGCCTTGGCCTCCCGCCCGGCGACGACCCCATGCTCCTCGCCTACCT
CCTGGACCCTTCCAACACCACCCCCGAGGGGGTGGCCCGGCGCTACGGCGGGGA
GTGGACGGAGGAGGCGGGGGAGCGGGCCGCCCTTTCCGAGAGGCTCTTCGCCAA
CCTGTGGGGGAGGCTTGAGGGGGAGGAGAGGCTCCTTTGGCTTTACCGGGAGGT
GGAGAGGCCCCTTTCCGCTGTCCTGGCCCACATGGAGGCCACGGGGGTGCGCCT
GGACGTGGCCTATCTCAGGGCCTTGTCCCTGGAGGTGGCCGAGGAGATCGCCCG
CCTCGAGGCCGAGGTCTTCCGCCTGGCCGGCCACCCCTTCAACCTCAACTCCCGG
GACCAGCTGGAAAGGGTCCTCTTTGACGAGCTAGGGCTTCCCGCCATCGGCAAG
ACGGAGAAGACCGGCAAGCGCTCCACCAGCGCCGCCGTCCTGGAGGCCCTCCGC
GAGGCCCACCCCATCGTGGAGAAGATCCTGCAGTACCGGGAGCTCACCAAGCTG
AAGAGCACCTACATTGACCCCTTGCCGGACCTCATCCACCCCAGGACGGGCCGCC
TCCACACCCGCTTCAACCAGACGGCCACGGCCACGGGCAGGCTAAGTAGCTCCG
ATCCCAACCTCCAGAACATCCCCGTCCGCACCCCGCTTGGGCAGAGGATCCGCCG
GGCCTTCATCGCCGAGGAGGGGTGGCTATTGGTGGCCCTGGACTATAGCCAGAT
AGAGCTCAGGGTGCTGGCCCACCTCTCCGGCGACGAGAACCTGATCCGGGTCTTC
CAGGAGGGGCGGGACATCCACACGGAGACCGCCAGCTGGATGTTCGGCGTCCCC
CGGGAGGCCGTGGACCCCCTGATGCGCCGGGCGGCCAAGACCATCAACTACGGG
GTCCTCTACGGCATGTCGGCCCACCGCCTCTCCCAGGAGCTAGCCATCCCTTACG
AGGAGGCCCAGGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCCAAGGTGCGGGC
CTGGATTGAGAAGACCCTGGAGGAGGGCAGGAGGCGGGGGTACGTGGAGACCC
TCTTCGGCCGCCGCCGCTACGTGCCAGACCTAGAGGCCCGGGTGAAGAGCGTGC
GGGAGGCGGCCGAGCGCATGGCCTTCAACATGCCCGTCCAGGGCACCGCCGCCG
ACCTCATGAAGCTGGCTATGGTGAAGCTCTTCCCCAGGCTGGAGGAAATGGGGG
CCAGGATGCTCCTTCAGGTCCACGACGAGCTGGTCCTCGAGGCCCCAAAAGAGA
GGGCGGAGGCCGTGGCCCGGCTGGCCAAGGAGGTCATGGAGGGGGTGTATCCCC
TGGCCGTGCCCCTGGAGGTGGAGGTGGGGATAGGGGAGGACTGGCTCTCCGCCA
AGGAGGGCATTGATGGCCGCGGCGGAGGCGGGCATCATCATCATCATCATTAA

SEQ ID NO:33 The amino acid sequence of Sso7d-ΔTaq (E5)(Y) protein

MITNSSATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGGGKTGRGAVSEKDA
PKELLQMLEKQKKGGGVTSGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGE

PVQAVYGFAKSLLKALKEDGDAVIVVFDAKAPSFPHEAYGGHKAGRAPTPEDFPRQ
LALIKELVDLLGLARLEVPGYEADDVLASLAKKAEKEGYEVRILTADKDLYQLLSDR
IHVLHPEGYLITPAWLWEKYGLRPDQWADYRALTGDESDNLPGVKGIGEKTARKLL
EEWGSLEALLKNLDRLKPAIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREP
DRERLRAFLERLEFGSLLHEFGLLESPKALEEAPWPPPEGAFVGFVLSRKEPMWADL
LALAAARGGRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLA
YLLDPSNTTPEGVARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREV
ERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQL
ERVLFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPL
PDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNIPVRTPLGQRIRRAFIAEEGWLLVA
LDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTI
NYGVLYGMSAHRLSQELAIPYEEAQAFIERYFQSFPKVRAWIEKTLEEGRRRGYVET
LFGRRRYVPDLEARVKSVREAAERMAFNMPVQGTAADLMKLAMVKLFPRLEEMG
ARMLLQVHDELVLEAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGIGEDWLSAK
EGIDGRGGGGHHHHHH

WHAT IS CLAIMED IS:

1. A method of increasing the yield from a polymerase reaction on a target nucleic acid present in a solution that comprises a polymerase inhibitor, the method comprising:

(a) contacting the target nucleic acid with a polymerase, wherein the polymerase is joined to a sequence-non-specific nucleic-acid-binding domain that: (i) binds to double-stranded nucleic acid, and (ii) enhances the processivity of the polymerase compared to an identical polymerase not having the sequence non-specific nucleic-acid-binding domain fused to it, and

wherein the solution is of a composition that permits the binding domain to bind to the target nucleic acid and the polymerase domain to extend a primer that is hybridized to the target nucleic acid sequence;

(b) incubating the solution under conditions in which the primer is extended by the polymerase.

2. A method of claim 1, wherein the polymerase inhibitor is a fluorescent dye.

3. A method of claim 2, wherein the dye is SYBR Green I.

4. A method of claim 1, wherein the polymerase domain has thermally stable polymerase activity.

5. A method of claim 4, wherein the thermally stable polymerase domain is a ΔTaq polymerase domain.

6. A method of claim 4, wherein the polymerase domain comprises a *Pyrococcus* polymerase domain.

7. A method of claim 1, wherein the sequence-non-specific nucleic-acid-binding domain specifically binds to polyclonal antibodies generated against Sso7d.

8. A method of claim 1, wherein the sequence-non-specific nucleic-acid-binding domain contains a 50 amino acid subsequence containing 50% amino acid similarity to Sso7d.

9. A method of claim 1, wherein the sequence-non-specific nucleic-acid-binding domain is Sso7d.

10. A method of claim 1, wherein the target nucleic acid is a crude sample.

11. A method of sequencing a nucleic acid in an aqueous solution using an improved polymerase, the method comprising:

(a) contacting the target nucleic acid with a polymerase, wherein the polymerase is joined to a sequence-non-specific nucleic-acid-binding domain that: (i) binds to double-stranded nucleic acid, and (ii) enhances the processivity of the polymerase compared to an identical polymerase not having the sequence non-specific nucleic-acid-binding domain fused to it, and

wherein the solution is of a composition that permits the binding domain to bind to the target nucleic acid and the polymerase domain to extend a primer that is hybridized to the target nucleic acid sequence;

(b) incubating the solution under conditions in which the primer is extended by the polymerase.

12. A method of claim 13, wherein the thermally stable polymerase domain is a ΔTaq polymerase domain.

13. A method of claim 11, wherein the sequence-non-specific nucleic-acid-binding domain specifically binds to polyclonal antibodies generated against Sso7d.

14. A method of claim 11, wherein the sequence-non-specific nucleic-acid-binding domain contains a 50 amino acid subsequence containing 50% amino acid similarity to Sso7d.

15. A method of claim 11, wherein the sequence-non-specific nucleic-acid-binding domain is Sso7d.

16. A method of performing a quantitative polymerase reaction on a target nucleic acid present in a solution that comprises a DNA-binding fluorescent dye, the method comprising: (a) contacting the target nucleic acid with a polymerase, wherein the polymerase is joined to a sequence-non-specific nucleic-acid-binding domain that: (i) binds to double-stranded nucleic acid, and (ii) enhances the processivity of the polymerase compared to an

identical polymerase not having the sequence non-specific nucleic-acid-binding domain fused to it, and

wherein the solution is of a composition that permits the binding domain to bind to the target nucleic acid and the polymerase domain to extend a primer that is hybridized to the target nucleic acid sequence;

(b) incubating the solution under conditions in which the primer is extended by the polymerase, and

(c) exposing the solution to a suitable excitation light and measuring fluorescence emission from the fluorescent dye.

17. A method of claim 16, wherein the polymerase domain has thermally stable polymerase activity.

18. A method of claim 17, wherein the thermally stable polymerase domain is a ΔTaq polymerase domain.

19. A method of claim 17, wherein the polymerase domain comprises a *Pyrococcus* polymerase domain.

20. A method of claim 16, wherein the sequence-non-specific nucleic-acid-binding domain specifically binds to polyclonal antibodies generated against Sso7d.

21. A method of claim 16, wherein the sequence-non-specific nucleic-acid-binding domain contains a 50 amino acid subsequence containing 50% amino acid similarity to Sso7d.

22. A method of claim 16, wherein the sequence-non-specific nucleic-acid-binding domain is Sso7d.

Figure 1A

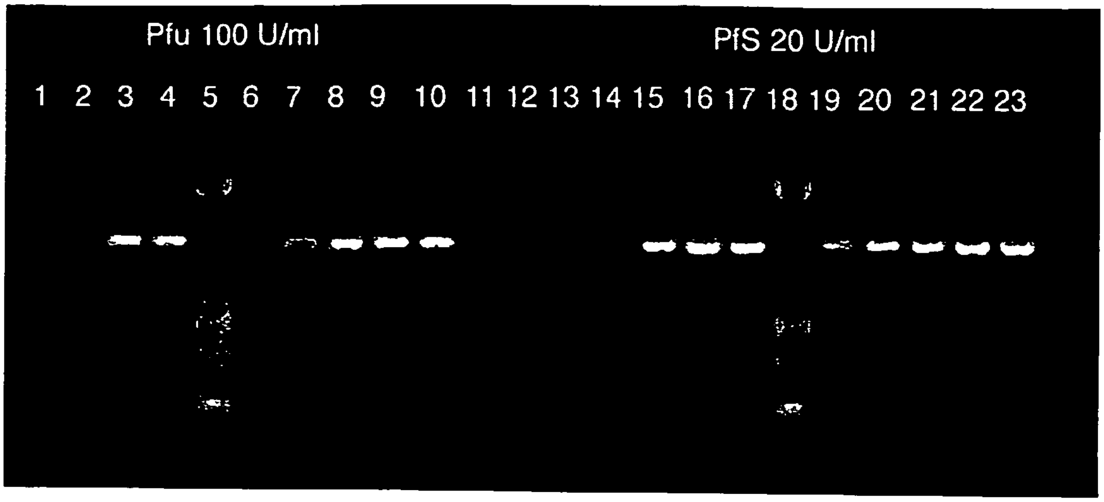

Pfu 100 U/ml
PfS 20 U/ml
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Figure 1B